CLAIMS

1.   A method of processing gene sequence data with use of one or more computers, the method comprising:

5   reading, by the computer, gene sequence data corresponding to a gene sequence and coding sequence data corresponding to a plurality of coding sequences within the gene sequence;

identifying, by the computer following a set of primer selection rules, primer pair data within the gene sequence data, the primer pair data corresponding to a pair of

10   primer sequences for one of the coding sequences, the set of primer selection rules including a first rule specifying that the primer pair data be obtained for a predetermined annealing temperature;

storing the primer pair data;

repeating the acts of identifying and storing such that primer pair data are

15   obtained for each sequence of the plurality of coding sequences at the predetermined annealing temperature; and

simultaneously amplifying the plurality of coding sequences in gene sequences from three or more individuals at the predetermined annealing temperature using the identified pairs of primer sequences, such that a plurality of amplified coding sequences

20   from the three or more individuals are obtained.

58

2. The method of claim 1, wherein the first rule further specifies that each primer sequence have a length that falls within one or more limited ranges of acceptable lengths.

3. The method of claim 1, wherein the set of primer selection rules includes a a second rule specifying that a single primer pair be identified for two or more coding regions if they are sufficiently close together.

4. The method of claim 1, wherein gene family data associated with the gene sequence is read by the computer, and the set of primer selection rules includes a second rule specifying that the primer pair data be excluded from the gene family data.

5. The method of claim 1, further comprising:

sequencing the plurality of amplified coding sequences to produce a plurality of nucleotide base identifier strings.

6. The method of claim 5, wherein the plurality of nucleotide base identifier strings includes nucleotide base identifiers represented by the letters G, A, T, and C.

7. The method of claim 6, further comprising:

positionally aligning, by the computer, the plurality of nucleotide base identifier strings to produce a plurality of aligned nucleotide base identifier strings.

59

8.   The method of claim 7, further comprising:

performing, by the computer, a comparison amongst aligned nucleotide base identifiers at each nucleotide base position of the plurality of aligned nucleotide base identifier strings.

9.   The method of claim 8, performing the following additional acts at each nucleotide base position where a difference amongst aligned nucleotide base identifiers exists:

reading, by the computer, nucleotide base quality information associated with the aligned nucleotide base identifiers where the difference exists;

comparing, by the computer, the nucleotide base quality information with predetermined qualification data;

visually displaying, from the computer, the nucleotide base quality information for acceptance or rejection; and

if the nucleotide base quality information meets the predetermined qualification data and is accepted: providing and storing resulting data that identifies where the difference amongst the aligned base identifiers exists.

10.   The method of claim 9, wherein the resulting data comprise single nucleotide polymorphism (SNP) identification data.

60

11.   The method of claim 9, wherein the nucleotide base quality information comprise one or more phred values.

12.   The method of claim 10, wherein after providing and storing all resulting data that identifies where the differences amongst the aligned nucleotide base identifiers exist, performing the following additional acts for each aligned nucleotide base identifier at each nucleotide base position where a difference exists:

comparing, by the computer, the nucleotide base identifier with a prestored nucleotide base identifier to identify whether the nucleotide base identifier is a variant; and

providing and storing, by the computer, additional resulting data that identifies whether the nucleotide base identifier is a variant.

13.   The method of claim 12, wherein the additional resulting data comprises haplotype identification data.

14.   The method of claim 13, wherein providing and storing additional resulting data comprises providing and storing a binary value of '0' for those nucleotide base identifiers that are identified as variants and a binary value of '1' for those nucleotide base identifiers that are not.

61

15.   A computer program product comprising:

a computer-usable storage medium;

computer-readable program code embodied on said computer-usable storage

medium; and

5   the computer-readable program code for effecting the following acts on a

computer:

reading gene sequence data corresponding to a gene sequence and coding

sequence data corresponding to a plurality of coding sequences within the gene

sequence;

10   identifying primer pair data within the gene sequence data by following a

set of primer selection rules, the primer pair data corresponding to a pair of

primer sequences for one of the coding sequences, the set of primer selection

rules including a first rule specifying that the primer pair data be obtained for a

predetermined annealing temperature;

15   storing the primer pair data;

repeating the acts of identifying and storing such that primer pair data are

obtained for each sequence of the plurality of coding sequences at the

predetermined annealing temperature, so that the plurality of coding sequences

can be simultaneously amplified in gene sequences from three or more of

20   individuals at the predetermined annealing temperature using the identified

pairs of primer sequences to produce a plurality of amplified coding sequences

from the three or more individuals.

62

16.   The computer program product of claim 15, wherein the first rule further specifies that each primer sequence have a length that falls within one or more limited ranges of acceptable lengths.

5

17.   The computer program product of claim 15, wherein the set of primer selection rules includes a second rule specifying that a single primer pair be identified for two or more coding regions if they are sufficiently close together.

10   18.   The computer program product of claim 15, wherein gene family data associated with the gene sequence is read by the computer, and the set of primer selection rules includes a second rule specifying that the primer sequence data be excluded from the gene family data.

15   19.   The computer program product of claim 15, wherein the plurality of amplified coding sequences are sequenced to produce a plurality of nucleotide base identifier strings.

20.   The computer program product of claim 19, wherein the plurality of

20   nucleotide base identifier strings includes nucleotide base identifiers represented by the letters G, A, T, and C.

21. The computer program product of claim 20, wherein the computer-readable program code is for effecting the following further acts on the computer:

positionally aligning the plurality of nucleotide base identifier strings to produce a plurality of aligned nucleotide base identifier strings.

5

22. The computer program product of claim 21, wherein the computer-readable program code is for effecting the following further acts on the computer:

performing a comparison amongst aligned nucleotide base identifiers at each nucleotide base position of the plurality of aligned nucleotide base identifier strings.

10

23. The computer program product of claim 22, wherein the computer-readable program code is for effecting the following additional acts at each nucleotide base position where a difference amongst aligned nucleotide base identifiers exists:

reading nucleotide base quality information associated with the aligned nucleotide base identifiers where the difference exists;

15

comparing the nucleotide base quality information with predetermined qualification data;

visually displaying the nucleotide base quality information for acceptance or rejection; and

20 if the nucleotide base quality information meets the predetermined qualification data and is accepted: providing and storing resulting data that identifies where the difference amongst the aligned base identifiers exists.

24.   The computer program product of claim 23, wherein the resulting data comprise single nucleotide polymorphism (SNP) identification data.

25.   The computer program product of claim 23, wherein the nucleotide base quality information comprise one or more phred values.

26.   The computer program product of claim 24, wherein after providing and storing all resulting data that identifies where the differences amongst the aligned nucleotide base identifiers exist, performing the following additional acts for each aligned nucleotide base identifier at each nucleotide base position where such difference exists:

comparing the nucleotide base identifier with a prestored nucleotide base identifier to identify whether the nucleotide base identifier is a variant; and

providing and storing additional resulting data that identifies whether the nucleotide base identifier is a variant.

27.   The computer program product of claim 26, wherein the additional resulting data comprises haplotype identification data.

28.   The computer program product of claim 27, wherein providing and storing additional resulting data comprises providing and storing a binary value of '0'

65

for those nucleotide base identifiers that are identified as variants and a binary value of '1' for those nucleotide base identifiers that are not.

29.   A method of processing gene sequence data with use of one or more computers, the method comprising:

reading, by the computer, a plurality of nucleotide base identifier strings;

positionally aligning, by the computer, the plurality of nucleotide base identifier strings to produce a plurality of aligned nucleotide base identifier strings;

performing, by the computer, a comparison amongst aligned nucleotide base identifiers at each nucleotide base position of the plurality of aligned nucleotide base identifier strings;

performing, by the computer, a comparison amongst aligned nucleotide base identifiers at each nucleotide base position of the plurality of aligned nucleotide base identifier strings;

at each nucleotide base position where a difference amongst aligned nucleotide base identifiers exists:

reading, by the computer, nucleotide base quality information associated with the aligned nucleotide base identifiers where the difference exists;

comparing, by the computer, the nucleotide base quality information with predetermined qualification data;

visually displaying, from the computer, the nucleotide base quality information for acceptance or rejection; and

66

if the nucleotide base quality information meets the predetermined qualification data and is accepted: providing and storing resulting data that identifies where the difference amongst the aligned base identifiers exists.

30. The method of claim 29, wherein the plurality of nucleotide base identifier strings includes nucleotide base identifiers represented by the letters G, A, T, and C.

31. The method of claim 30, wherein the resulting data comprise single nucleotide polymorphism (SNP) identification data.

32. The method of claim 31, wherein the nucleotide base quality information comprise one or more phred values.

33. The method of claim 31, wherein after providing and storing all resulting data that identifies where the differences amongst the aligned nucleotide base identifiers exist, performing the following additional acts for each aligned nucleotide base identifier at each nucleotide base position where such difference exists:

comparing, by the computer, the nucleotide base identifier with a prestored nucleotide base identifier to identify whether the nucleotide base identifier is a variant; and

providing and storing, by the computer, additional resulting data that identifies whether the nucleotide base identifier is a variant.

67

34. The method of claim 33, wherein the additional resulting data comprises haplotype identification data.

35. The method of claim 34, wherein providing and storing additional resulting data comprises providing and storing a binary value of '0' for those nucleotide base identifiers that are identified as variants and a binary value of '1' for those nucleotide base identifiers that are not.

36. A computer program product comprising:

a computer-usable storage medium;

computer-readable program code embodied on said computer-usable storage medium; and

the computer-readable program code for effecting the following acts on a computer:

reading a plurality of nucleotide base identifier strings;

positionally aligning the plurality of nucleotide base identifier strings to produce a plurality of aligned nucleotide base identifier strings;

performing a comparison amongst aligned nucleotide base identifiers at each nucleotide base position of the plurality of aligned nucleotide base identifier strings;

68

performing a comparison amongst aligned nucleotide base identifiers at each nucleotide base position of the plurality of aligned nucleotide base identifier strings;

at each nucleotide base position where a difference amongst aligned

5   nucleotide base identifiers exists:

reading nucleotide base quality information associated with the aligned nucleotide base identifiers where the difference exists;

comparing the nucleotide base quality information with predetermined qualification data;

10   visually displaying the nucleotide base quality information for acceptance or rejection; and

if the nucleotide base quality information meets the predetermined qualification data and is accepted: providing and storing resulting data that identifies where the difference amongst the aligned base identifiers

15   exists.

37.   The computer program product of claim 36, wherein the plurality of nucleotide base identifier strings includes nucleotide base identifiers represented by the letters G, A, T, and C.

20

38.   The computer program product of claim 37, wherein the resulting data comprise single nucleotide polymorphism (SNP) identification data.

69

39.   The computer program product of claim 38, wherein the nucleotide base quality information comprise one or more phred values.

5   40.   The computer program product of claim 38, wherein after providing and storing resulting data that identifies where the differences amongst the aligned nucleotide base identifiers exist, performing the following additional acts for each aligned nucleotide base identifier at each nucleotide base position where such difference exists:

10   comparing the nucleotide base identifier with a prestored nucleotide base identifier to identify whether the nucleotide base identifier is a variant; and

providing and storing additional resulting data that identifies whether the nucleotide base identifier is a variant.

15   41.   The computer program product of claim 40, wherein the additional resulting data comprises haplotype identification data.

42.   The computer program product of claim 41, wherein providing and storing additional resulting data comprises providing and storing a binary value of '0'

20   for those nucleotide base identifiers that are identified as variants and a binary value of '1' for those nucleotide base identifiers that are not.

43.   A method of processing gene sequence data with use of one or more computers, the method comprising:

reading, by the computer, gene sequence data corresponding to a gene sequence and coding sequence data corresponding to a plurality of coding sequences within the

5   gene sequence;

identifying, by the computer following a set of primer selection rules, primer pair data within the gene sequence data, the primer pair data corresponding to a pair of primer sequences for one of the coding sequences, the set of primer selection rules including a first rule specifying that the primer pair data be obtained for a

10   predetermined annealing temperature and a second rule specifying that a single primer pair be identified for two or more coding regions if they are sufficiently close together;

storing, by the computer, the primer pair data; and

repeating the acts of identifying and storing such that primer pair data are obtained for the plurality of coding sequences at the predetermined annealing

15   temperature.

44.   The method of claim 43, further comprising:

simultaneously amplifying the plurality of coding sequences in gene sequences from three or more of individuals at the predetermined annealing temperature using

20   the identified pairs of primer sequences, so that a plurality of amplified coding sequences from the three or more individuals are obtained.

45. The method of claim 43, wherein gene family data associated with the gene sequence is read by the computer, and the set of primer selection rules includes a third rule specifying that the primer sequence data be excluded from the gene family data.

5